

# Two Concepts of Metacognition

**Peter Carruthers**

**Abstract:** This comment distinguishes two concepts of metacognition that are often conflated by Smith et al. (this volume). One refers to any executively controlled process, the other to metarepresentational processes. It is argued that more progress will be made by comparative researchers if they target their investigations separately at these two phenomena, as well as studying the relationships between them.

**Keywords:** executive, mental rehearsal, metacognition, metarepresentation, working memory

I endorse the critique provided by Smith et al. (this volume) of associationist explanations of the primate metacognition data. I also agree that an obsessive focus on associationist accounts of animal behavior impedes progress in comparative psychology and obstructs attempts to understand animal precursors and homologies of components of human cognition. However, more care needs to be taken by Smith and colleagues (in this and previous publications) to draw distinctions among potential non-associationist explanations of their data. In particular, they use the term “metacognition” to describe two sets of processes that are really quite different. This, too, has the effect of impeding progress in comparative psychology, and gets in the way of understanding the commonalities and differences among the cognitive processes of humans and other animals.

In one sense, a metacognitive process is any process that is “meta” to or “above” other cognitive processes by monitoring and controlling them (but without metarepresenting them). In this sense, all executive processes are metacognitive, especially those that direct attention, resolve conflicts, activate explicit long-term memories, and control and manipulate the contents of working memory. In another sense, metacognitive processes are those that involve self-directed metarepresentational states (in contrast with other-directed metarepresentational states, which are employed when reading the minds of other agents). Processes of this latter sort are the ones generally studied in the human metacognition literature, which frequently examines metarepresentational reports of people’s feelings of knowing, degrees of confidence, and so

forth. To avoid confusion, for present purposes I shall drop all use of the terms “metacognition” and “metacognitive”, and would urge Smith and colleagues to do the same. Instead, I shall speak of *executive* processes and *metarepresentational* processes respectively.

Notice that in humans many executive processes (especially those that implicate working memory) are accompanied by some degree of metarepresentation. This is because the contents of working memory are “globally broadcast” (in the sense of Baars, 1988) to a wide range of other systems in the brain (just as attended perceptual contents are). Included among such systems will be the human mindreading system, or whatever other system is responsible for self-directed metarepresentation (Carruthers, 2011). Hence a long-term memory that is activated and sustained by attention in working memory, for example, may routinely be categorized by humans *as* a memory, and will be reported by people as such. It is far from clear, however, that metarepresentational classifications are always necessary for memories to do their work. Indeed, there is evidence that they are not, since people suffering from autism, who often have severe difficulties with metarepresentation generally, can nevertheless activate and deploy memories successfully (Minshew & Goldstein, 2001). While metarepresentational processes no doubt play important roles in many forms of executive function, they are by no means essential to all executive activities.

The data reviewed by Smith et al. (this volume) provide strong support for executively controlled processes in nonhuman primates. Consider the memory-monitoring experiments, for example. In order to choose adaptively when deciding whether or not to take the test, the animal either needs to have sustained a working-memory image of the target stimulus, or it needs to activate such an image from long-term memory. Provided a sufficiently vivid image is present it should take the test, otherwise it should decline. Yet the capacity to sustain a representation in working memory is known to depend on executively directed attention in humans and other animals, and the mechanisms of such attention are largely homologous across primate species, at least (Carruthers, in press). Similarly, there is now considerable evidence of episodic-like long-term memory systems that are homologous across all mammals, as well as mechanisms that are at least partly homologous in birds (Allen & Fortin, in press). The activation of such memories is likewise dependent on executively controlled attentional search.

Note that a human might explain her strategy in taking one of these memory-monitoring tests by saying, “If I could recall what was on the screen initially, then I took the test; otherwise I

opted out.” While this is a metarepresentational report, there is little reason to think that the strategy, when implemented, is itself metarepresentational. Rather, both humans and other animals can operate with the decision-making schema, “Take the test only if X was on the screen initially [where X is the content of a memory image]; otherwise decline.” Similarly, a human might provide a metarepresentational report of her attempt to probe long-term memory by saying, “I was trying to remember what had been on the screen.” But again, the process of memory search itself is unlikely to be metarepresentational. Rather, both humans and other animals need to direct first-order questions to long-term memory systems, asking themselves (as it were), “What was on the screen initially?” The answer (if one is found) will be a first-order representation that can activate the decision-making schema above.

Similar points can be made about the uncertainty monitoring data reviewed by Smith et al. (this volume). Humans who participate in such experiments generally explain their choices in metarepresentational terms, saying, for example, “I opted out because I felt uncertain whether the stimulus was sparse or dense.” But again, while such choices are executively controlled, it is far from clear that they result from processes that are themselves metarepresentational. Rather, as Carruthers & Ritchie (2012) argue, they may result from affective responses caused by executively controlled rehearsal of the available actions in working memory. (Notice that this would explain Smith et al.’s [in press] finding that uncertainty responding is disrupted by concurrent attentional load.) Indeed, the literature on human decision making suggests that this is what standardly happens (Damasio, 1994; Gilbert & Wilson, 2007). Rehearsal of an action (such as pressing the *dense* key) leads to a working memory representation of such an action being executed. In cases where judgment is difficult, the action will be appraised as unlikely to succeed, given one’s background knowledge of the contingencies of the experiment. As a result, negative valence is created by one’s valuational systems and directed at the represented action, causing it to seem like a bad alternative. Nothing metarepresentational needs to take place.

Notice that this account makes a prediction, which to my knowledge has not yet been tested. This is that one ought to be able to find a dissociation between metarepresentational judgments of uncertainty and opting-out performance in populations of people whose metarepresentational capacities are poor (such as those suffering from autism or schizophrenia). To the extent that basic executive and working memory capacities remain intact in such people, we should expect that they will continue to opt out adaptively in conditions of uncertainty. But

with metarepresentational capacities damaged, their judgments of their own uncertainty should correlate much less well with objective difficulty, as well as with their actual performance.

Indeed, once the two concepts of metacognition distinguished here are separated cleanly from one another in the way that I have urged, one can envisage a productive comparative research program that explores capacities for executive control and the circumstances in which such control is, or is not, dependent on or influenced by metarepresentational processes, in both humans and other animals. Smith et al. (this volume) should follow their own advice: they should stop focusing exclusively on the basic contrast between associative and non-associative processes, and begin to explore different forms of the latter on a comparative basis.

## References

- Allen, T. & Fortin, N. (in press). Evolution of episodic memory. *Proceedings of the National Academy of Sciences*.
- Baars, B. (1988). *A Cognitive Theory of Consciousness*. NY: Cambridge University Press.
- Carruthers, P. (2011). *The Opacity of Mind*. NY: Oxford University Press.
- Carruthers, P. (in press). Evolution of working memory. *Proceedings of the National Academy of Sciences*.
- Carruthers, P., & Ritchie, J. B. (2012). Metacognition and uncertainty in animals. In Beran, M., Brandl, J., Perner, J., & Proust, J. (Eds.), *The Foundations of Metacognition*. Oxford: Oxford University Press.
- Damasio, A. (1994). *Descartes' Error*. London: Papermac.
- Gilbert, D. & Wilson, T. (2007). Prospection: Experiencing the future. *Science*, 317, 1351-1354.
- Minschew, N. & Goldstein, G. (2001). The pattern of intact and impaired memory functions in autism. *Journal of Child Psychology and Psychiatry*, 42, 1095-1101.
- Smith, J. D., Couchman, J., & Beran, M. (this volume). Animal metacognition: A tale of two comparative psychologies.
- Smith, J. D., Coutinho, M., Church, B., & Beran, M. (in press). Executive-attentional uncertainty responses by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: General*.